Journal of Forestry Research (2013) 24(2): 227–236
DOI 10.1007/s11676-012-0302-0

ORIGINAL PAPER

Analysis of genetic relationship among *Arbutus unedo* L. genotypes using RAPD and SSR markers

Filomena Gomes • Rita Costa • Maria M. Ribeiro
Elisa Figueiredo •Jorge M. Canhoto

Received: 2011-12-19; Accepted: 2012-01-26

Abstract: The strawberry tree (*Arbutus unedo* L.) is an underutilized, drought tolerant, fire resistant species with a south western distribution in Europe, and with ecological and putative socio-economical impact in Portugal and Mediterranean countries. Our aim was to develop an appropriate set of molecular markers to enable genetic diversity to be assessed and to fingerprint *Arbutus unedo* genotypes for breeding and conservation purposes in Portugal. Twenty-seven trees from a broad geographic range were screened with 20 random amplified polymorphic DNA (RAPD primers) and 11 microsatellite markers (SSR). The RAPDs generated 124 bands, 57.3% of which were polymorphic, with an expected heterozygosity of 27%. We cross-amplified 11 SSR primers developed for *Vaccinium* spp., and 5 were found to be polymorphic in *A. unedo*, with 75% of expected heterozygosity, a number of alleles of 11.6, a null allele frequency of 7.6% and a polymorphic information content of 71%. Although the SSRs were more polymorphic and informative than the RAPDs, both markers displayed high genetic variability with the gathered data. No geographic pattern was observed in the genetic variation distribution based on both marker systems, and the lack of correlation between genetic and geographical matrices was confirmed by Mantel tests. Likely, no correlation was found between pairwise SSR and RAPD band-sharing matrices. These results and their implications on *A. unedo* breeding and conservation programs are discussed.

Keywords: Ericaceae; fingerprinting; geographic pattern; molecular markers; strawberry tree.

The online version is available at http:// www.springerlink.com

Filomena Gomes (✉)
CERNAS, Departamento Florestal, Escola Superior Agrária Coimbra, Bencanta, 3040-316, Coimbra, Portugal, Tel: 351 962.650.955, Fax: 351 239.802.979, Email: fgomes@esac.pt

Rita Costa
INRB, Instituto Nacional de Recursos Biológicos, IP/L-INIA, Av. República, Quinta do Marquês 2780-159 Oeiras, Portugal.

Maria M. Ribeiro
Departamento de Recursos Naturais e Desenvolvimento Sustentável, Escola Superior Agrária, 6001-909 Castelo Branco. Portugal

Elisa Figueiredo • Jorge M. Canhoto
Centre of Functional Ecology, Department of Life Sciences, University of Coimbra, Ap. 3046, 3001-401 Coimbra, Portugal

Responsible editor: Chai Ruihai

Introduction

The strawberry tree (*Arbutus unedo* L.) is an evergreen shrub or small tree belonging to the Ericaceae family with a circum-Mediterranean distribution, growing in regions where temperatures are amenable (Torres et al. 2002). According to the International Centre for Underutilized Crops (www.cropsforthefuture.org) and the Global Facilitation Unit for Underutilized Species (www.underutilized-species.org) this species falls into the category of neglected or underutilized crops. Therefore, it is an undervalued fruit tree, with different possible commercial uses from processed and fresh fruit production to ornamental, pharmaceutical and chemical industrial applications, due to the phenolic acids and terpenoid compounds with strong antioxidant activity, vitamin C and tannin content (Celikel et al. 2008). In addition, it is fire resistant and, due to its pioneer status, it is valuable for land recovery and desertification avoidance (Piotto et al. 2001).

In Portugal *A. unedo* is widely distributed, from Atlantic climate areas in the North to dry arid areas in the South, occupying about 15,500 ha (Godinho-Ferreira et al. 2005) and, to our knowledge, its genetic diversity status is unknown. According to Pedro (1994) *A. unedo* rarely constitutes dominant stands being more common in patchy bush-like communities or in natural stands dominated by oaks. The species appears naturally in different phytosociological alliances, from the cork-oak woodlands (*Sanguisorbo-Quercetum suberis*) and the strawberry-tree dominated scrub (*Arbuto-Cistetum populifolii*) in the South, to the pedunculate oak-woodlands (*Rusceto-Quercetum roboris*) in the North, including the oak-woodlands (*Arisaro-Quercetum bro-*

teroi and *Arbuto-Quercetum pyrenaicae*) and strawberry dominated scrub (*Phillyreo-Arbutetum unedonis*) in the Centre of the country. *Arbutus unedo* extensive areas, though patchy, occur mainly in the Southern mountainous regions (Serra de Monchique and Caldeirão), whereas a more fragmented distribution is common in central and northern regions, due to intensive forestation programs with *Pinus pinaster* and *Eucalyptus globulus* that form closed canopy stands. Noteworthy, in some areas, *A. unedo* spread due to recurrent fires, ecosystem degradation and abandoned farmland, growing in large patches of shrub-like formations (Meireles et al. 2005). Indeed, the traditional human forest management since medieval ages favoured the distribution patterns of several insect-pollinated, light-demanding European species, which were originally scattered in the natural landscape (Wohlgemuth et al. 2002). Fruits are usually collected from spontaneous field-growing genotypes for liqueur production, which represents the foremost income. Germplasm diversity evaluation is needed for plant breeding and conservation purposes. Provenance or common garden experiments provide information about adaptive traits under natural selection, unlike molecular markers which are, usually, selectively neutral. Nevertheless, molecular data, besides fingerprinting, provides information about the possible material origin, gene flow and genetic diversity. PCR-based molecular markers have been widely used in genetic diversity and fingerprinting studies (Bassil et al. 2006). The random amplified polymorphic DNA (RAPD) markers rely on the use of short PCR primers, and do not require sequence information or laborious cloning. Therefore, the method's speed, sensitivity and versatility make it suitable for a rapid polymorphism survey (Spooner et al. 2005). Despite RAPD markers limitations (dominance and protocol sensitiveness), they have been used to fingerprint and to evaluate genetic diversity in several species, due to their simplicity and applicability. RAPDs proved to be a valuable tool to evaluate genetic diversity in *A. unedo* (Takrouni and Boussaid 2010), in other Ericaceae, such as *Rhododendron* spp. (Milne and Abbott 2008) and *Vaccinium* spp. (e.g. Albert et al. 2005; Debnath 2007) and, also, to detect genetic relationships among accessions, e. g. *Vaccinium myrtillus* (Albert et al. 2003).

Microsatellites (SSR) are currently used for fingerprinting, inbreeding and genetic structure studies, because of their high polymorphism, co-dominance, multiallelism, automation analysis and ubiquitous occurrence (Ellegren 2004). These markers have, also, been successfully used to screen polymorphism in members of the Ericaceae family, such as *Rhododendron* spp. (Tan et al. 2009; Wang et al. 2010) and *Vaccinium* spp. (Bassil et al. 2006; Bassil et al. 2010; Boches et al. 2005; Debnath 2010; Hirai et al. 2010). Their discovery is an expensive and time-consuming process, but among closely related species cross-transferability is often successful. The *Vaccinium* genus comprises over 400 species, and due to the commercial importance of blueberries (section *Cyanococcus*) and cranberries (section *Oxycoccus*), molecular markers, particularly SSR, have been developed to assess genetic diversity and to fingerprint germplasm collections (e.g. Bassil et al. 2006; Boches et al. 2005). According to Boches et al. (2005), the ability to cross amplify SSR *loci* in a panel of 12 species, representing different sections of the *Vaccinium* genus were, on average, 83% successful. *Arbutus unedo* is included in the Arbutoideae, a distinct and natural group within the Ericaceae, among the earlier *taxa* in the evolution of Ericaceae, which branches to a clade that contains the Vaccinioideae (Hileman et al. 2001; Kron et al. 2002). Thus, based on the studies performed in *Vaccinium* (Bassil et al. 2006; Boches et al. 2005), we selected and tested 11 SSRs in *A. unedo*. To our knowledge, strawberry tree fingerprinting and genetic diversity studies based on SSR markers were not yet been made.

The main objectives of this work were to test 20 RAPD primers and to cross-amplify 11 microsatellites from *Vaccinium* species in *A. unedo* genotypes: (1) to develop a set of SSR and RAPD markers to fingerprint strawberry tree genotypes selected across the species distribution range in Portugal, which have been evaluated for biotechnological purposes (Gomes and Canhoto 2009; Gomes et al. 2010), (2) to assess its genetic diversity, and (3) to discuss the utility of this set of markers for breeding and conservation purposes.

Material and methods

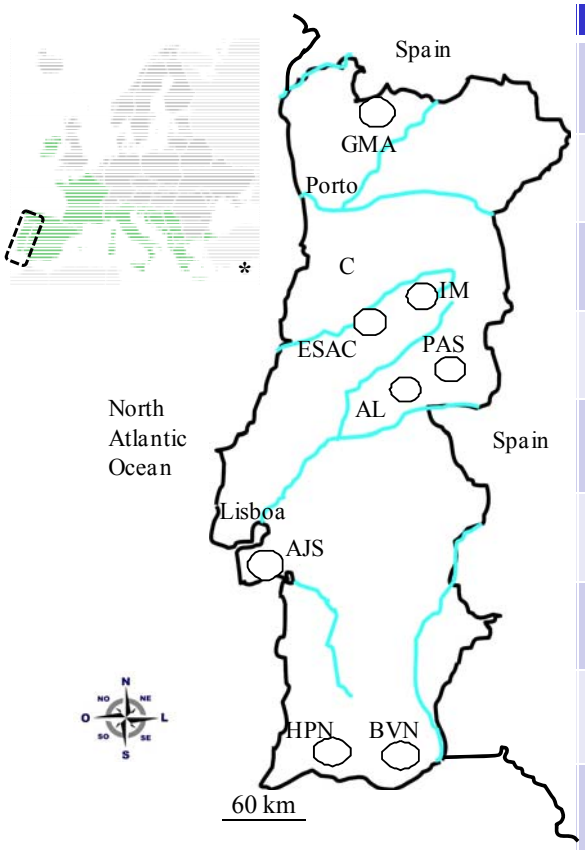
Plant material and DNA extraction

The sampling strategy was to collect plant material of *A. unedo* from three trees in each of the nine regions of Portugal where there is a history of human usage of the species for many years (Fig. 1). In two regions (AL and IM, Fig. 1) the plant material was collected from trees that were phenotypically superior in terms of fruit production and quality. In the remaining regions, samples were taken from natural stands where no information about productivity was available. The twenty-seven trees were collected in the following way (Fig. 1): three in the north, growing on granitic soils (GMA); six in the centre (C; ESAC); six accessions were selected from trees growing on schist-derived soils (IM and AL, respectively); three growing on schist-derived soils, in the east (PAS); three growing on calcareous soils, in the west (AJS) and six in the south (three from BVN and three from HPN), growing also, on schist-derived soils. Eleven of those genotypes, selected for fruit production and quality, were established *in vitro* and in clonal tests (Au2, Au3, Au5, Au9, Au13, Au16, Au17, Au18, Au24, Au27 and Au28) for future breeding purposes and to provide the stakeholders with improved plant material (Gomes and Canhoto 2009; Gomes et al. 2010). Young leaves were harvested from each tree, frozen in liquid nitrogen, and stored at -80°C. The DNA was extracted using the DNeasy Plant Mini kit (Qiagen).

RAPDs genotyping

Twenty arbitrary primers, decamer oligonucleotides from Operon Technologies Kit C (OPC, Table 1) were tested. For each primer the PCR reactions were prepared as master mixes to minimize errors. Each PCR reaction contained 2.5 mM $MgCl_2$, 0.2 mM of each dNTPs, 2 μM of primer, 0.5 U of Taq DNA

polymerase in 1 X reaction buffer and 50 ng of genomic DNA in a final 13.5-µl volume. The initial denaturation step (5 min, 94ºC) was followed by 35 cycles of 30 s at 94ºC, 45 s at 35ºC and, finally, 72°C for 10 min. The PCR reaction products were separated by electrophoresis in agarose gel (1.5%) with Sybr safe DNA (1.5µl/100 ml), and visualized under UV light, together with fragment size standard. The PCR reactions were repeated at least three times in order to confirm the reproducibility of the results.



Regions	Code	Location
GMA	Au 7	Gerês, Mata de Albergaria
	Au 10	
	Au 22	
C	Au 3	Coimbra
	Au 11	
	Au 15	
ESAC	Au 21	Coimbra
	Au 23	
	Au 27	
IM	Au 13	Serra do Açor
	Au 16	
	Au 24	
PAS	Au 6	Serra da Gardunha
	Au 9	
	Au 19	
AL	Au 2	Serra Alvélos
	Au 5	
	Au 17	
AJS	Au 1	Serra da Arrábida
	Au 20	
	Au 25	
HPN	Au 4	Algarve, São Marcos da Serra,
	Au 8	
	Au 28	
BVN	Au 12	Algarve, Serra do Caldeirão
	Au 18	
	Au 26	

Fig. 1 The map indicates the provenances of the sampled material from different regions of Portugal (circles), whereas the chart gives the particular locations where the plant material was collected and respective code.

SSRs genotyping

Based on the previous works of Boches et al. (2005) and Bassil et al. (2006), 11 SSRs were selected on the basis of detected polymorphism, the number of alleles per locus, the allele scoring quality, and the repeat unit (Table 2).

The PCR reactions were used a master mix, as described in the above section. The amplification conditions followed the protocol described by Boches et al. (2005) and optimized for *A. unedo*. The optimum annealing temperature (TaºC) for each primer was determined by PCR gradient from 55ºC to 65ºC. After the initial denaturation step at 94ºC for 4 min, DNA was amplified for 35 to 40 cycles in a thermocycler programmed for 45s denaturation step at 94ºC, a 45 s annealing step at the optimum annealing temperature of the primer pair, a 45s extension step at 72ºC, and a final extension step at 72ºC for 30 min in a total of 6 steps. Since the primers CA169F and NA741 showed multiple peaks, a touch-down for these primers was tested in PCR reactions with 9 steps instead of 6. Two TaºC for 30 s were performed (62ºC and 57ºC for 10 and 30 cycles, respectively). The PCR reactions were repeated at least four times in order to confirm the reproducibility of the results. For allele scoring and sizing, fluorescently labelled forward primers (FAM, HEX, or NED) and unlabelled reverse primers were used. The PCR products were diluted with HI-DI formamide for denaturation, and ROX marker was added. The amplified products were denatured and run on an automatic sequencer (ABI 310 Applied Biosystems) and the results were scored using Genescan software (Applied Biosystems).

Diversity estimates

RAPDs: The 20 RAPD primers, from Operon Technologies Kit C (OPC) fragments obtained from the 27 genotypes were scored in the form of a binary matrix (1/0 = presence/ absence of a

band). This matrix was used to evaluate pairwise genetic similarity, calculated with the Lynch similarity coefficient (Lynch 1990), which is a band-sharing based method. An UPGMA-based dendrogram was constructed using the NTSYS-PC 2.02i software (Rohlf 1997). The reliability of the generated dendrogram was tested through a Mantel test by using bootstrap analysis with 1000 permutations. The Mantel test statistic was also used to measure the degree of relationship between the geographic and the genetic distance using the same software.

The genetic diversity analysis was performed with the AFLP-SURV 1.0 software (Vekemans et al. 2002). The diversity parameters comprised the number of polymorphic loci (P) and the expected heterozygosity (He) (Nei 1987). A matrix of the geographic distances between every two individuals was compared with the genetic similarity distance matrix known as the Lynch coefficient (Lynch 1990) with a Mantel test, as described above.

SSRs: The diversity parameters and the polymorphic information content (PIC) were investigated using the Cervus 3.0 software (Marshall et al. 1998). The polymorphism information content (PIC) for each SSR primer was determined separately using the following equation:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - 2\left[\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} p_i^2 p_j^2\right],$$

where p_i is the frequency of the i^{th} allele, and n is the number of alleles (Botstein et al. 1980).

The diversity parameters comprised the number of alleles (Na), the observed heterozygosity (Ho), the expected heterozygosity (He) (Nei 1987), and the fixation index (F_{is}) (Weir and Cockerham 1984). The Hardy-Weinberg equilibrium (HWE) was tested for each locus (Markov-Chain method), the null allele frequencies (F *Null*) per loci were estimated using a maximum likelihood EM algorithm and the linkage disequilibria (LD) tests were performed for all loci combinations. All these analyses were performed with the *Genepop* software. Genetic similarity was assessed using the band-sharing coefficient described above. The UPGMA tree topology was tested by comparing the similarity pairwise matrix and the correspondent cophenetic matrix through a Mantel test (Sokal 1979). This test was also used to measure the degree of relationship between geographic and genetic distances and to further test the relationships between the two molecular markers matrices.

Results

Random amplified polymorphic DNA (RAPD)

Nineteen out of the 20 tested RAPD primer pairs produced amplified products, and 16 had polymorphic profiles. Fig 2 illustrates the amplification products of the primer OPC-10. The 19 primers generated a total of 124 bands, ranging from 200 to 2100 bp, and 57.3% were polymorphic (71 loci) (Table 1). The expected heterozygosity (He) was 0.27±0.014.

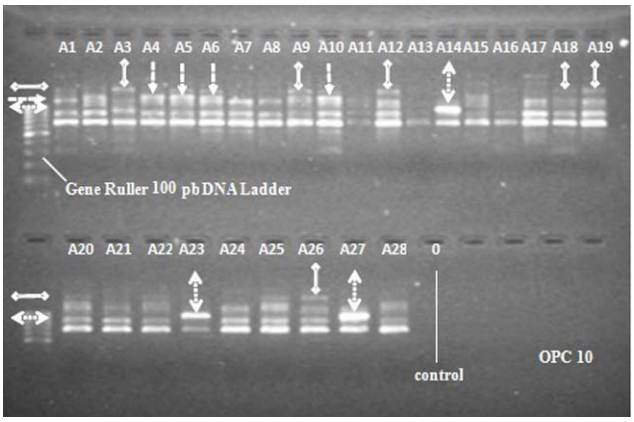


Fig. 2 Agarose gel amplified products from OPC-10 RAPD marker after electrophoresis and visualized under UV light. The *A. unedo* genotypes are referred as A1 to A28, instead of Au1 to Au28, for easier reading (A14 and A23 are samples of the same genotype and 0 is the control without DNA). Polymorphisms are indicated by arrows, each type corresponding to a band size from the ladder in the leftmost lane.

Table 1. Information obtained with the 20 OPC primers used in RAPD analysis of *A. unedo*

Primer (OPC)	Number of bands	Range size (bp)	Polymorphic bands Number (%)
OPC 1	7	500 - 2000	4 (57.1 %)
OPC 2	6	200 - 1200	2 (33.3 %)
OPC 3	8	400 - 1500	5 (62.5 %)
OPC 4	9	400 - 2000	5 (55.6 %)
OPC 5	6	300 - 1500	3 (50.0 %)
OPC 6	9	300 - 2100	6 (66.7 %)
OPC 7	4	400 - 1450	2 (50.0 %)
OPC 8	4	400 -1200	0 (0.0 %)
OPC 9	5	650 - 1700	2 (40.0 %)
OPC 10	10	300 - 1500	8 (80.0 %)
OPC 11	13	220 - 1600	12 (92.3 %)
OPC 12	8	400 - 2000	6 (75.0 %)
OPC 13	9	500 - 1700	4 (44.4 %)
OPC 14	5	650 - 2100	1 (20.0 %)
OPC 15	0	-	-
OPC 16	7	400 - 1900	5 (71.4 %)
OPC 17	1	1600	0 (0 %)
OPC 18	5	300 - 1800	3 (60.0 %)
OPC 19	7	540 - 1900	3 (42.9 %)
OPC 20	1	400	0 (0 %)
Total	**124**	**200 – 2100**	**71 (57.3%)**

The cluster analysis based in the Lynch coefficient revealed an among genotypes similarity of 83% (Fig. 3a). Additionally, some genotypes shared as much as 95% of the bands. The Mantel test confirmed the UPGMA topology, with a moderate yet significant correlation (matrix correlation: r=0.64; P<0.001).

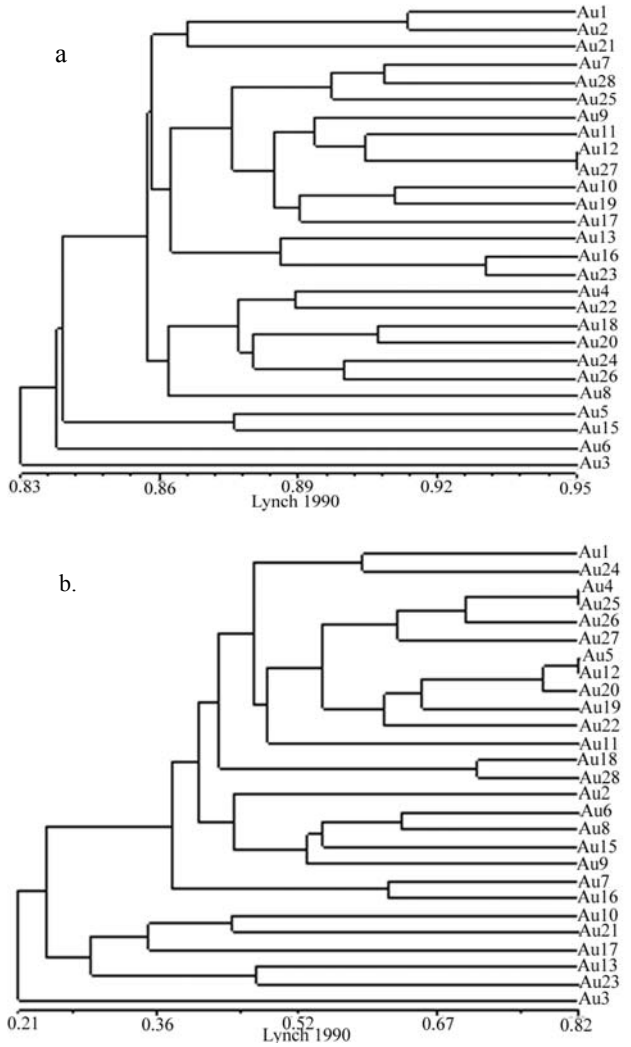


Fig. 3 a. Genetic similarity analysis of 27 *A. unedo* genotypes using the Lynch (1990) coefficient based on 19 of the 20 RAPDs primers that showed amplified PCR products. b. Genetic similarity analysis of 27 *A. unedo* genotypes using the same coefficient based on 5 SSRs polymorphic *loci*.

The Mantel test also confirmed that there was no correlation between genetic and geographical distances matrices (matrix correlation: r=0.01; p<0.57), so genotypes could not be grouped according to their geographical origin. Indeed a consistent geographic pattern among genotypes could not be found as can be seen by comparing Au12 and Au27 (from the BVN and the ESAC regions, respectively; Fig. 1).

Microsatellites (SSRs)

Only two primers (CA855F and VCC_K4) out of the 11 that were tested failed to cross-amplify in *A. unedo*, and both had the biggest repeat motif (compound repeat: $(GA)_{14}(CGA)_5$ and $(TC)_{16}(TC)_{12}$, Table 2). Three loci were monomorphic and one locus (NA1040) was discarded due to complex banding pattern (Table 2). Partial amplified products, referring to few individuals, from two SSR *loci* (CA421F and NA398), visualized and scored with automatic sequencer and Genescan software, are displayed in Fig. 4.

Table 2. Microsatellite informativeness based on the repeat motif, the optimum annealing temperature (TaºC), the existence of amplification products (+/-), polymorphism (+/-), complex banding pattern (C), and allele size range (bp).

SSR	Repeat motif	TaºC	Amplification products	Polymorphism	Allele size range (bp)
CA169F	$(GAT)_4$	62-57	+	-	103-114
CA421F	$(CT)_{25}$	60	+	+	150-211
CA855F	$(GA)_{14}(CGA)_5$	X	-	-	---
NA800	$(TC)_{13}$	58	+	+	169-278
NA398	$(AAAT)_5$	55	+	-	210
VCC_K4	$(TC)_{16}(TC)_{12}$	X	-	-	---
CA794F	$(GA)_{12}$	60	+	+	283-294
NA961	$(TAC)_5$	60	+	-	186-208
NA1040	$(TC)_{11}$	60	+	C	171-239
NA741	$(TC)_9$	62-57	+	+	165-172
VCC_I2	$(CT)_{14}$	60	+	+	174-274

The five polymorphic loci had a mean expected and observed heterozygosity of 0.75 and 0.70, respectively. The mean number of alleles (Na) ranged from 6 to 22 per locus with a mean across the five loci of 11.6 (Table 3). The observed heterozygosity was lower than the expected heterozygosity for all loci, indicating a putative heterozygote deficit. Indeed a mean homozygote excess was found (7%), albeit not significant (Table 3). The results showed that 4 out of the 5 polymorphic SSR loci were in HWE equilibrium (after Bonferroni correction). Additionally, the null allele frequency (F *Null*) was about 8% on average (Table 3), which may explain the observed heterozygote deficiency (though the individuals do not constitute a population and are not assumed to be in HWE). The mean polymorphism information content (PIC) was 0.71. Four loci showed PIC values higher than 0.70, thereby indicating that they should be useful diversity indicators. The locus CA421F displayed values higher than the average: 22 alleles, He=0.90 and a high PIC value (88%). Conversely, locus NA741 showed the lowest PIC value (51%), number of alleles (6), and expected diversity (0.57). The linkage *disequilibrium* test showed that all the loci are independent.

Table 3. Diversity parameters obtained with the 5 SSR polymorphic loci after screening the *A. unedo* accessions.

SSR	Na	He	Ho	Fis	P-value	Sig.	F (*Null*)	PIC
CA421F	22	0.90	0.85	0.05	0.2018	NS	0.0546	0.88
NA800	9	0.79	0.78	0.02	0.0146	NS	0.0534	0.75
CA794F	7	0.78	0.69	0.11	0.0159	NS	0.0985	0.73
NA741	6	0.57	0.56	0.03	0.004	NS	0.0868	0.51
VCC_I2	14	0.73	0.62	0.16	0.0007	*	0.0887	0.70
Mean	11.6	0.75	0.70	0.07	0.0474		0.0764	0.71

Na = number of alleles per locus; He = expected heterozygosity; Ho = observed heterozygosity; Fis = fixation index; PIC = polymorphic information content; *Sig.* = significance resulting from the HWE test (after Bonferroni correction); NS - not significant and *- significant; *Null* = null allele frequency estimates.

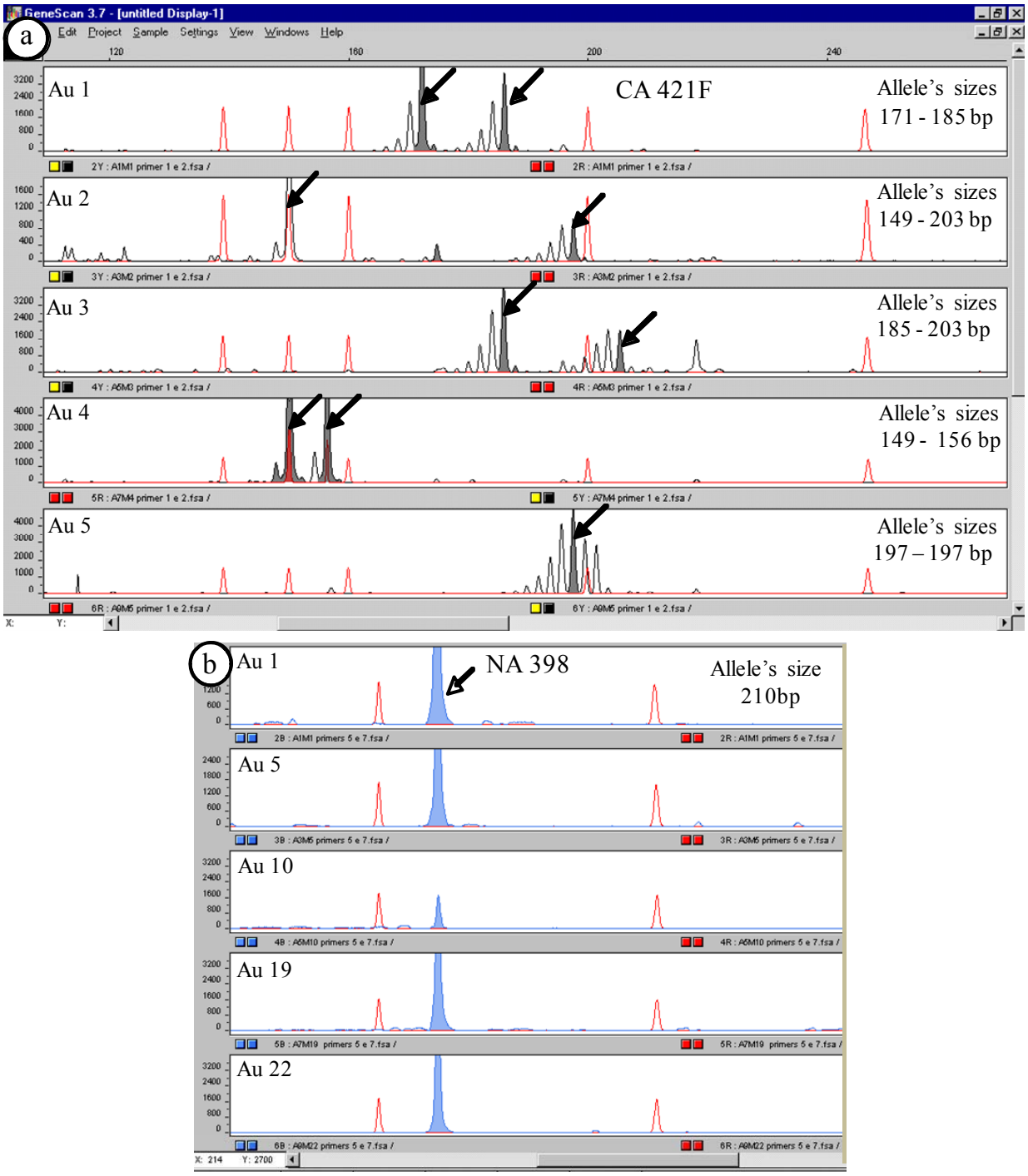


Fig. 4 The amplified products from SSRs, the filled arrows point to the band profiles: **a. Polymorphic primer (CA421F). b. Monomorphic primer (NA398)**. Visualized in an automatic sequencer (ABI 310 Applied Biosystems) and scored using Genescan software (Applied Biosystems). The genotype code and the allele's size (bp) are noted. Red peak refers to the ROX marker.

Pairwise genetic similarity analysis using the Lynch coefficient indicated that the level of allele sharing between genotypes ranged from 21% to 82% (Fig. 3b). The Mantel test confirmed the tree topology (r=0.75; p<0.001). According to the Lynch coefficient, the genotypes Au4 and Au25 (from HPN and AJS provenances), and Au5 and Au12 (from AL and BVN provenances) shared as much as 82% alleles. However, these genotypes did not show a consistent pattern of clustering according to their geographic origins (see Fig. 1). No genotypes could be grouped according to their geographical origin, which was confirmed by the Mantel test (matrix correlation: r=0.09; p<0.17). This test also confirmed the absence of correlation between pairwise molecular markers (RAPDs and SSRs) matrices (matrix correlation: r=0.05; p<0.69).

Discussion

Markers' informativeness

In the literature, the reported polymorphism using RAPD markers to screen plant genotypes were, in some cases, similar to the value we obtained in the current study (57%). For example, Agrama and Tuinstra (2003) found a similar value, 60% of polymorphic bands, using 32 RAPDs primers to screen 22 sorghum genotypes. A slightly lower diversity was observed in Tunisian *A. unedo* populations (He = 0.216), in this case the deforestation followed by species fragmentation and consequent bottleneck and genetic drift were linked to the lower diversity levels found (Takrouni and Boussaid 2010). In other reports, the estimated values were much higher or lower. Majourhat et al. (2008) obtained higher number of polymorphic bands in *Argania spinosa* (~96%) and Schneller et al. (1998) estimated 18% of polymorphic bands in *Dryopteris remota*, using 22 genotypes and 19 RAPD primers. According to the authors, the lower diversity unfold in *D. remota* may be due to putative unique geographic origin and apomictic status. These differences may be explained by the species' own diversity, its fragmentation, a more or less extended plant material distribution area used in the studies, the species' mating system and gene flow, and the use of natural or artificial regeneration together with the plant propagation method (seedlings or clones).

SSRs are particularly attractive for distinguishing among genotypes since they are co-dominant and have a higher level of polymorphism than other molecular markers (Ellegren, 2004). As a short-cut, we cross-amplified 11 SSR loci, and the five loci we used displayed a relevant heterozygosity (Ho= 0.70), negligible presence of null alleles (F null= 0.076), 11.6 number of alleles per locus and a 0.71 informativeness mean (PIC). The locus NA741 showed the poorest informative results with a PIC value bellow 70%, and the remaining loci were highly informative (PIC≥0.70). Locus informativeness showed to be more correlated with He than to Na. For example, locus VCC_12 had 70% of PIC and 14 alleles, but locus NA800 with only 9 alleles had a higher PIC (0.75), in accordance with the results from a study by Fernandes et al. (2008) where three highly informative loci were used to fingerprint *Pinus pinaster* genotypes. Interestingly, the results we obtained showed that the primer (NA741) with the lowest PIC value (0.51) had also the lowest values of Na (6), Ho (0.56) and He (0.57). Conversely, locus CA421F, with the highest PIC value (88%), displayed the highest values for genetic diversity parameters: Na=22, Ho= 0.85, He=0.90 and PIC=0.88.

Avramidou et al. (2010) obtained much lower average PIC value (0.473) in a similar study, where 36 wild cherry (*Prunus avium*) genotypes selected in nine populations were screened with 14 SSR, compared to the value we obtained. This value reflected the low genetic diversity in the genotypes assembly and the authors concluded that those results mirror the material they used (a relatively small assembly of individuals collected in different populations). Nevertheless, in the *Pinus pinaster* study reported above, where 60 plus trees were genotyped with 3 SSR, the average He value obtained was slightly higher than the value obtained in that species' population genetic study (Derory et al. 2002). The melting pot of individuals selected in different populations might result in a higher genetic diversity value, since more alleles are sampled, unless the populations have very similar allelic composition. Probably the fact that the reproduction in wild cherry also occurs asexually via suckering might reflect a generally lower values of within population diversity and number of alleles (Vaughan et al. 2007), and individuals, particularly in unmanaged populations, are likely to be closely related to or of the same clone as their nearest neighbors (< 100 m).

In the current study, with the SSRs, we have found that the expected heterozygosity (He=0.75) was slightly higher than the observed heterozygosity (Ho=0.70), further reflected on a 7%, fixation index, on average, yet not significant. Positive and significant Fis values mirror differences between observed and expected heterozygosity, due to putative heterozygosity loss as a result of non-random mating of parents. It must be emphasized, though, that the trees used in our experiments were collected in different stands, producing an assembly unlikely in the Hardy-Weinberg equilibrium, as previously referred. Anyway, only one locus (VCC_I2) displayed significant deviation from what the Hardy-Weinberg equilibrium (HWE) would anticipate.

The presence of null alleles is another factor that may contribute to the deviation from HWE. Indeed, the null allele frequency estimates was about 7.6%, on average. Noteworthy, the only locus with significant deviation from HWE had similar null allele frequency estimate compared to the other loci (0.089), but a sounding Fis value (16%). Tan et al. (2009) using 20 individuals from two *Rhododendron simsii* populations, reported that 3 out of the 8 SSR loci they used showed significant deviation from HWE, after applying Bonferroni correction, but in this case with higher null alleles estimates, ranging from 15% to 44%.

Microsatellite transferability

Cross amplification is a powerful tool to circumvent the cost and time requested to develop microsatellites *de novo*. Although a low transferability of genomic SSRs in related genera has been reported in many species (e.g. Gupta and Varshney 2000; Mariette et al. 2001), including Ericaceae (Kameyama et al. 2006); only one of the 13 SSR primers developed for *Rhododendron metternichii* cross amplified in *Phyllodoce aleutica* and *P. caerulea*), successful cross amplification was reported in several species, including Ericaceae. Indeed, ten microsatellites were successfully transferred from *Prunus persica* to fingerprint elite wild cherry genotypes (Avramidou et al. 2010). In Ericaceae, 11 polymorphic SSRs were developed and revealed polymorphism in *Monotropa hypopitys*, and a subset of those primers successfully amplified in the congener *M. uniflora* and in five other closely related genera (Klooster et al. 2008). Indeed, Bassil et al. (2010) cross-amplified 18 SSR isolated from domestic *Vaccinium* in *V. reticulatum, V. calycinum, V. myrtillus* and in other endemic species, and those SSR revealed to be polymorphic in almost all the evaluated genotypes. In the current study, as previously re-

ferred, we successfully cross-amplified nine of the primers developed by Boches et al. (2005) and Bassil et al. (2006) in *Vaccinium*, and five loci showed to be polymorphic and informative.

Marker diversity and clustering comparison

The SSR markers revealed higher levels of genetic polymorphism than the RAPD markers in the current study. SSRs' expected heterozygosity value, 75%, was considerably higher than the 27% expected with the RAPDs. This observation is consistent with the higher levels of polymorphism usually displayed by SSRs, thus the results are in agreement with germplasm analysis studies carried out in other species (e.g. Agrama and Tuinstra 2003; Powell et al. 1996; Rajora and Rahman 2003; Sorkheh et al. 2009). Moreover, the pairwise band-sharing range was broader in the SSRs (0 and 0.82) than in the RAPDs (0.77 and 0.95). Rajora and Rahman (2003) also obtained larger range with SSR compared with RAPD in 17 *Populus* cultivars, and Agrama and Tuinstra (2003) reported that the average genetic similarity (simple matching coefficient) between sorghum genotypes was lower when estimated with the SSR markers (0.44) compared to RAPDs (0.61). Additionally, closely related persimmon genotypes shared the same RAPD marker profile making difficult to distinguish among them (Badenes et al. 2003), whereas the SSR analysis proved to be useful to differentiate several genotypes (Naval et al. 2010). Also, in our experiments a lack of correlation was found between SSR and RAPD pairwise similarity matrices, was probably due to the higher SSR ability to discriminate genotypes compared to RAPDs.

In the current study the physical distance between every two genotypes compared to the genetic distance was very low and not significant, using both markers. The average distance of the selected trees is ca. 262 Km, and any genetic pattern should only hold true at close vicinity, probably less than a few tens of meters. For instance, with the SSRs, the tree Au13 and Au7 which grow 271 km apart, share 0% of the bands (the minimum pairwise Lynch value) whereas, trees Au4 and Au25, which are 235 km apart, share 83% of the bands (the maximum value). The same pictures is obtained with RAPDs, Au3 and Au6 160 km distant, have 77% of common bands (the minimum), and Au12 apart 420 km from Au27 share 95% of the bands (the maximum value). This lack of correlation was also verified by Takrouni and Boussaid (2010) in *A. unedo*, using RAPDs and Majourhat et al. (2008) also uncovered no geographical pattern of genetic variation in *Argania spinosa*. Conversely, using 28 microsatellite markers Bassil et al. (2006) demonstrated a geographic pattern to the distribution of diversity in *V. corymbosum* 69 wild and domesticated accessions.

In the current study it is possible to suggest some reasons for the observed lack of geographic/genetic correlation. These include i) differences between pollen and seed flow (seed dispersal by birds and pollen flow by small insects) fostered by the low number of the screened genotypes sampled spaced out from each other, ii) species fragmentation, iii) mating system, and iv) putative human-mediated species dissemination. *Arbutus* flowers are pollinated by bees and other insects and the seeds are dispersed by birds, rodents, and other animals, but also by gravity (see below) (Beland et al. 2005). *Arbutus* has a patchy or fragmentary distribution and, for instance, bees that pollinate the trees tend to forage in an area with a radius less than 500 m (Dramstad 1996). By one hand, the low gene flow at small scale would prevent exchange of genetic material between relatively close trees (500 m), and, on the other hand, birds can transport large quantities of seeds over long distances, thereby connecting distant populations (Jordano et al. 2007). Thus, trees separated by hundreds of kilometres could share bands. The same kind of results were obtained in argan tree (*Argania spinosa*) by El-Mousadik and Petit (1996). The authors found a high level of gene flow by seed and relatively low level of gene flow by pollen. In this case goats and camels were responsible for nut dispersal up to at least few tens of kilometres. In another study concerning the mating system in *Arbutus menziesii*, the authors referred a very high outcrossing rate, circa 97% (Beland et al. 2005). They observed fruits beneath and around the trees in the field, demonstrating that *Arbutus* seeds are besides bird- are also gravity-dispersed (which is also quite common in *A. unedo*). Consequently, in *Arbutus menziesii* kinship decreases exponentially with distance between trees due, possibly, to open pollinated progenies established close to maternal trees, and the populations are well structured at small scale, but exhibited no geographic trends at large scale.

Additionally, *A. unedo* dispersal by anthropogenic activities should not be excluded, since *A. unedo* has long been used for fruit consumption and fire use, among others. The human impact in the Mediterranean region, blurring the plants' geographic genetic pattern is well known, as in the case of *Pinus pinaster* in Portugal (Ribeiro et al. 2001). The same situation, the lack of geographical pattern, was not found in 45 wild cranberry (*V. macrocarpon*), possibly reflecting the high rates of gene flow (by seeds) due both to human migration and agricultural trade (Debnath 2007).

Finally, it must be stressed that through the markers we have used in *A. unedo* all the screened genotypes were uniquely fingerprinted, an aspect of the utmost importance for germplasm conservation and tree fingerprinting. The results of this study will allow the development of new strategies for managing strawberry tree germplasm. The set of markers we developed will also be useful in future studies, such as the analysis of the species' genetic structure, to check genetic uniformity of cloned plants (*in vitro* or *ex vitro*), to uncover variability for breeding purposes and to develop marker-assisted selection systems, among others, with an increased number of individuals selected in populations within provenances. Additionally, germplasm conservation strategies in this species could be based on gathered information with the aim at preserving biodiversity and to initiate a conservation program.

Acknowledgments

We thank to all the people who helped us to obtain the plant material. Thanks are extended to Isabel Evaristo for help in the laboratory work and to Stephanie Mariette for reviewing the manuscript. F. Gomes was supported by a PhD fellowship (SFRH/BD/37170/2007) from the Portuguese Foundation for

Science and Technology (FCT).

References

Agrama HA, Tuinstra MR. 2003. Phylogenetic diversity and relationships among sorghum accessions using SSRs and RAPDs. *African Journal of Biotechnology*, **2**: 334–340.

Albert T, Raspe O, Jacquemart AL. 2003. Clonal structure in *Vaccinium myrtillus* L. revealed by RAPD and AFLP markers. *International Journal of Plant Sciences*, **164**: 649–655.

Albert T, Raspe O, Jacquemart AL. 2005. Diversity and spatial structure of clones in *Vaccinium uliginosum* populations. *Canadian Journal of Botany*, **83:** 211–218.

Avramidou E, Ganopoulos IV, Aravanopoulos FA. 2010. DNA fingerprinting of elite Greek wild cherry. *Forestry*, **83**: 527–533.

Badenes M, Garcés A, Romero C, Romero M, Clavé J, Rovira M, Llácer G. 2003. Genetic diversity of introduced and local Spanish persimmon cultivars revealed by RAPD markers. *Genetic Resources and Crop Evolution*, **50:** 579–585.

Bassil NV, Boches P, Rowland L. 2006. Genetic diversity in the highbush blueberry evaluated with microsatellite markers. *Journal of the American Society for Horticultural Science*, **131**: 674–686.

Bassil NV, Bunch T, Nyberg A, Hummer K, Zee FT. 2010. Microsatellite markers distinguish Hawailan ohelo from other *Vaccinium* L. section *Myrtillus* species. *Acta Horticulturae*, **859:** 81–88.

Beland JD, Krakowski J, Ritland CE, Ritland K, El-kassaby YA. 2005. Genetic structure and mating system of northern *Arbutus menziesii* (Ericaceae) populations. *Canadian Journal of Botany*, **83:** 1581–1589.

Boches PS, Bassil NV, Rowland LJ. 2005. Microsatellite markers for *Vaccinium* from EST and genomic library. *Molecular Ecology*, **5**: 657–660.

Botstein WRL, Skolnick K, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics*, **32**: 314–331.

Celikel G, Demirsoy L, Demirsoy H. 2008. The strawberry tree (*Arbutus unedo* L.) selection in Turkey. *Scientia Horticulturae,* **118**: 115–119.

Debnath SC. 2007. An assessment of the genetic diversity within a collection of wild Cranberry (V*accinium macrocarpon* Ait.) clones with RAPD-PCR. *Genetic Resources and Crop Evolution*, **54**: 509–517.

Debnath SC. 2010. Berry bioreactor micropropagation and molecular analysis for genetic fidelity. *Acta Horticulturae*, **865**: 43–51.

Derory J, Mariette S, Gonzaléz-Martínez SC, Chagné D, Madur D, Gerber S, Brach J, Persyn F, Ribeiro MM, Plomion C. 2002. What can nuclear microsatellites tell us about maritime pine genetic resources conservation and provenance certification strategies? *Annals of Forest Science*, **59:** 699–708.

Dramstad W. 1996. Do bumblebees (*Hymenoptera*: Apidae) really forage close to their nests? *Journal of Insect Behavior*, **9**: 163–182.

El-Mousadik A, Petit RJ. 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology*, **5**: 547-555.

Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, **5**: 435–445.

Fernandes L, Rocheta M, Cordeiro J, Pereira S, Gerber S, Oliveira, MM, Ribeiro MM. 2008. Genetic variation, matting patterns and gene flow in a *Pinus pinaster* Aiton clonal seed orchard. *Annals of Forest Science*, **65:** 1–9.

Godinho-Ferreira PG, Azevedo AM, Rego F. 2005. Carta da tipologia florestal de Portugal Continental. *Silva Lusitana*, **13**: 1–34.

Gomes F, Canhoto JM. 2009. Micropropagation of strawberry tree (*Arbutus unedo* L.) from adult plants. *In Vitro Cellular & Developmental Biology - Plant*, **45**: 72–82.

Gomes F, Simões M, Lopes ML, Canhoto, JM. 2010. Effect of plant growth regulators and genotype on the micropropagation of adult trees of *Arbutus unedo* L. (strawberry tree). *New Biotechnology*, **27**: 882–892.

Gupta PK, Varshney RK. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with special emphasis on bread wheat. *Euphytica*, **113**: 163–185.

Hileman LC, Vasey MC, Thomas Parker V. 2001. Phylogeny and Biogeography of the Arbutoideae (Ericaceae): Implications for the Madrean-Tethyan Hypothesis. *Systematic Botany,* **26**: 131–143.

Hirai M, Yoshimura S, Ohsako T, Kubo N. 2010. Genetic diversity and phylogenetic relationships of the endangered species *Vaccinium sieboldii* and *Vaccinium ciliatum* (Ericaceae). *Plant Systematics and Evolution*, **287:** 75–84.

Jordano P, García C, Godoy JA, García-Castaño JL. 2007. Differential contribution of frugivores to complex seed dispersal patterns. *Proceedings National Academy of Sciences USA*, **104:** 3278–3282.

Kameyama Y, Kasagi T, Kudo, G. 2006. Eight microsatellite markers for sympatric alpine shrubs, *Phyllodoce aleutica* and *P. caerulea* (Ericaceae). *Molecular Ecology*, **6**: 402–404.

Klooster MR, Hoenle AW, Culley TM. 2008. Characterization of microsatellite *loci* in the myco-heterotrophic plant *Monotropa hypopitys* (Ericaceae) and amplification in related taxa. *Molecular Ecology Resources*, **9**: 219–221.

Kron KA, Stevens PF, Crayn DM, Anderberg AA, Gadek PA, Quinn CJ, Luteyn JL. 2002. Phylogenetic classification of Ericaceae: molecular and morphological evidence. *The Botanical Review*, **68:** 335–423.

Lynch M. 1990. The similarity index and DNA fingerprinting. *Molecular Biology Evolution*, **7**: 478–484.

Majourhat K, Jabba Y, Hafidi A, Martínez-Gómez P. 2008. Molecular characterization and genetic relationships among most common identified morphotypes of critically endangered rare Moroccan species *Argania spinosa* (Sapotaceae) using RAPD and SSR markers. *Annals of Forest Science*, **65**: 1–6.

Mariette S, Chagne D, Decroocq S, Vendramin GG, Lalanne C, Madur D, Plomion C. 2001. Microsatellite markers for *Pinus pinaster* Ait. *Annals of Forest Science*, **58**: 203–206.

Marshall TC, Slate J, Kruuk L, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**: 639–655.

Meireles C, Gonçalves P, Rego F, Silveira S. 2005. Estudo da regeneração natural das espécies arbóreas autóctones na Reserva Natural da Serra da Malcata. *Silva Lusitana*, **13**: 217 – 231.

Milne RI, Abbott RJ. 2008. Reproductive isolation among two interfertile *Rhododendron* species: low frequency of post-F-1 hybrid genotypes in alpine hybrid zones. *Molecular Ecology*, **17:** 1108–1121.

Naval M, Zuriaga E, Pecchioli S, Llácer G, Giordani E, Badenes M. 2010. Analysis of genetic diversity among persimmon cultivars using microsatellite markers. *Tree Genetics & Genomes*, **6**: 677–687.

Nei M. 1987. *Molecular evolutionary genetics*. New York: Columbia University Press, p.512.

Pedro JG. 1994. *Portugal Atlas do Ambiente. Notícia Explicativa II.6 Carta da distribuição de figueira e medronheiro*. Lisboa: Ministério do Ambiente e Recursos Naturais, p.39.

Piotto B, Piccini C, Arcadu P. 2001. La ripresa della vegetazione dopo gli incendi nella regione mediterrânea. In: B. Piotto and A. Noi (eds.), *Propagazione per seme di alberi e arbusti della flora mediterranea*. Roma: Dipartimento Prevenzione e Risanamento Ambientali, pp. 32–38.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, **2**: 225–238.

Rajora OP, Rahman MH. 2003. Microsatellite DNA and RAPD fingerprinting, identification and genetic relationships of hybrid poplar (*Populus* x *canadensis*) cultivars. *Theoretical and Applied Genetics*, **106**: 470 –477.

Ribeiro MM, Plomion C, Petit R, Vendramin GG, Szmidt AE. 2001. Variation in chloroplast single-sequence repeats in Portuguese maritime pine (*Pinus pinaster* Ait.). *Theoretical and Applied Genetics*, **102**: 97–103.

Rohlf FJ. 1997. *NTSYS-pc. Numerical taxonomy and multivariate analysis system version 2.01*. New York: Dep. of Ecology and Evolution, State University of New York, p.38.

Schneller J, Holderegger R, Gugerli F, Eichenberger K, Lutz E. 1998. Patterns of genetic variation detected by RAPDs suggest a single origin with subsequent mutations and long-distance dispersal in the apomictic fern *Dryopteris remota* (Dryopteridaceae). *American Journal of Botany*, **85:** 1038–1042.

Sokal R. 1979. Testing statistical significance of geographic variation patterns. *Systematic Zoology*, **28**: 227–232.

Sorkheh K, Shiran B, Kiani S, Amirbakhtiar N, Mousavi S, Rouhi V, Mohammady-D S, Gradziel TM, Malysheva-Otto LV, Martínez-Gómez P. 2009. Discriminating ability of molecular markers and morphological characterization in the establishment of genetic relationships in cultivated genotypes of almond and related wild species. *Journal Forestry Research*, **20:** 183−194.

Spooner D, Van Treuren R, Vicente MC. 2005. *Molecular Markers for Genebank Management.* Rome: International Plant Genetic Resources Institute (IPGRI), Technical Bulletin Nº 10.

Takrouni MM, Boussaid M. 2010. Genetic diversity and population's structure in Tunisian strawberry tree (*Arbutus unedo* L.). *Scientia Horticulturae*, **126**: 330–337.

Tan X-X, Li Y, Ge X-J. 2009. Development and characterization of eight polymorphic microsatellites for *Rhododendron simsii* Planch (Ericaceae). *Conservation Genetics*, **10**: 1553–1555.

Torres JA, Valle F, Pinto C, Garcia-Fuentes A, Salazar C, Cano E. 2002. *Arbutus unedo* communities in southern Iberian Peninsula mountains. *Plant Ecology*, **160**: 207–223.

Vaughan SP, Cottrell JE, Moodley DJ, Connolly T, Russell1 K. 2007. Distribution and fine-scale spatial-genetic structure in British wild cherry (*Prunus avium* L.). *Heredity*, **98**: 274–283.

Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I. 2002. Data from amplified fragment polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, **11**: 139–151.

Wang N, Qin ZC, Yang JB, Zhang JL. 2010. Development and characterization of 15 microsatellite *loci* for *Rhododendron delavayi* Franch. (Ericaceae). *Hortscience*, **45**: 457–459.

Weir CC, Cockerham BS. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, **38**: 1358–1370.

Wohlgemuth T, Bürgi M, Scheidegger C, Schütz M. 2002. Dominance reduction of species through disturbance - a proposed management principle for Central European forests. *Forest Ecology and Management*: **166**, 1–15.